



Published in final edited form as:

J Glaucoma. 2018 September ; 27(9): 785–793. doi:10.1097/IJG.0000000000001005.

Prediction Accuracy of the Dynamic Structure-Function model for Glaucoma Progression using Contrast Sensitivity Perimetry and Confocal Scanning Laser Ophthalmoscopy

Koosha Ramezani, MD^{1,2}, Iván Marín-Franch, PhD¹, Rongrong Hu, MD, PhD^{2,3}, William H Swanson, PhD⁴, and Lyne Racette, PhD^{1,2}

¹Department of Ophthalmology and Visual Sciences, School of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA

²Eugene and Marilyn Glick Eye Institute, Department of Ophthalmology, Indiana University School of Medicine, Indianapolis, IN, USA

³First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China

⁴Indiana University School of Optometry, Bloomington, IN, USA

Abstract

Purpose—The purpose of this study was to determine whether combining a structural measure with contrast sensitivity perimetry (CSP), which has lower test-retest variability than static automated perimetry (SAP), reduces prediction error with two models of glaucoma progression.

Methods—In this retrospective analysis, eyes with 5 visits with rim area (RA), SAP and CSP measures were selected from two datasets. Twenty-six eyes with open-angle glaucoma were included in the analyses. For CSP and SAP, mean sensitivity (MS) was obtained by converting the sensitivity values at each location from decibel (SAP) or log units (CSP) to linear units, and then averaging all values. MS and RA values were expressed as percent of mean normal based on independent normative data. Data from the first 3 and 4 visits were used to calculate errors in prediction for the 4th and 5th visits, respectively. Prediction errors were obtained for simple linear regression and the dynamic structure-function (DSF) model.

Results—With linear regression, the median prediction errors ranged from 6% to 17% when SAP MS and RA were used and from 9% to 17% when CSP MS and RA were used. With the DSF model, the median prediction errors ranged from 6% to 11% when SAP MS and RA were used and from 7% to 16% when CSP MS and RA were used.

Conclusions—The DSF model had consistently lower prediction errors than simple linear regression. The lower test-retest variability of CSP in glaucomatous defects did not, however, result in lower prediction error.

Correspondence to: Lyne Racette, PhD, Department of Ophthalmology and Visual Sciences, School of Medicine, University of Alabama at Birmingham, 1720 University Blvd, Suite 609, Birmingham, AL, 35294-009, Phone: 205-325-8673, lracette@uabmc.edu.

Conflict of interests: Dr. William H Swanson is an unpaid consultant for Heidelberg Engineering (Heidelberg, Germany). None of the other co-authors have any conflict of interests.

Keywords

Glaucoma; structure-function; perimetry; contrast sensitivity

INTRODUCTION

Glaucoma is a group of optic neuropathies in which progressive degeneration of the retinal ganglion cells and their axons results in visual field loss.¹ Detecting the earliest signs of glaucomatous progression is important in clinical practice in order to modify or intensify the treatment of glaucoma to prevent irreversible pathologic changes.² Progression of the disease can be detected by observing structural, as well as functional change.³ The nature and mechanism of the relationship between structure and function in glaucoma is not yet fully understood.⁴ Structural and functional measurements may each provide unique information regarding the presence of the disease and its progression.⁵

Attempts have been made to improve sensitivity to detect glaucoma progression. Several models such as machine learning classifiers and Bayesian models have been developed to identify progression based on combined structural and functional measurements.^{6,7} The Bayesian approach has been shown to provide significantly better performance for detecting glaucoma progression compared to ordinary least square linear regression.^{8,9} The dynamic structure-function model (DSF) is another model developed to monitor progression that uses structural and functional data jointly.¹⁰ We previously showed that the DSF model has lower prediction error compared to ordinary linear regression when short follow-up series are used. These results were obtained using rim area (RA) as the structural parameter and mean of sensitivity values expressed in linear units (MS) obtained from static automated perimetry (SAP) as the functional parameter.

SAP is widely used for visual field assessment and is the current clinical standard. Nevertheless, its high test-retest variability in areas of glaucomatous defects poses a challenge when monitoring glaucoma progression.¹¹ For SAP, it has been calculated that a reduction of about 20% in test-retest variability is necessary to detect progression one visit earlier and that a reduction of about 40% is necessary to detect progression a year earlier.¹²

In part to address the drawbacks of SAP and reduce its variability, other tests of visual function have been developed using different types of stimuli^{13,14} or stimulus of different sizes.¹⁵ Contrast sensitivity perimetry (CSP) is one such test that uses sinusoidal stimuli to achieve low variability in areas of glaucomatous defects. CSP uses sinusoidal stimuli with peak spatial frequencies similar to those used in frequency-doubling technology perimetry (FDP), but without the sharp stimulus edges and high temporal frequencies used in FDP.¹⁴ CSP has been shown to have lower test-retest variability in damaged regions of the visual field, as well as better agreement with RA compared to SAP.¹⁴ This may lead to earlier detection of glaucoma progression. The aim of this study was to assess the prediction accuracy of the DSF model when CSP is used as a functional test compared to SAP. Given the lower test-retest variability of CSP compared to SAP, we hypothesized that using it as the functional test would improve the prediction accuracy of the DSF model.

METHODS

Participants

This study was a retrospective data analysis based on the two datasets reported in Swanson et al and Hot et al.^{14,16} From these datasets, we selected patients with glaucoma who had at least six visits with HRT, SAP and CSP-1 data (n=19) and HRT, SAP and CSP-2 data (n=20). Some patients were included in both datasets (n=13) (a total of 26 eyes of 26 patients were selected across the two datasets). Patients were recruited from Indiana University School of Optometry and from the State University of New York College of Optometry. The 19 patients tested with CSP-1 had mean (\pm SD) age of 66 ± 7 years of age at baseline. The 20 patients tested with CSP-2 had mean (\pm SD) age of 69 ± 7 years of age at baseline. Overall, 46% of the patients were females. The study adhered to the tenets of the declaration of Helsinki, was HIPAA-compliant and approved by all appropriate institutional review boards. Informed consent was obtained from each participant after the goals and procedures of the study were explained.

Inclusion criteria—All patients had a diagnosis of glaucoma, clear ocular media, spherical correction within 6 diopters (D) and cylindrical correction within 3 D, intra-ocular pressure below 22 mmHg under treatment, and visual field defects consistent with a diagnosis of glaucoma. The diagnosis of glaucoma was established based on medical and family history, slit lamp biomicroscopy, applanation tonometry, detailed funduscopy, stereoscopic ophthalmoscopy of the optic disc with a 78-D lens, stereophotographs of the optic nerve, optic nerve imaging, and visual field evaluation.

Exclusion criteria—Patients were excluded from the study if they had an eye disease other than glaucoma, or if they had a systemic disease known to affect vision, or if they were taking medications known to affect visual function.

Structural measurements

Confocal scanning laser ophthalmoscopy scans of the optic nerve head were acquired with the Heidelberg Retina Tomograph II (HRT II) (Heidelberg Engineering, Heidelberg, Germany). These scans provided quantitative analysis of the neuroretinal RA. The quality of the image was evaluated by the HRT software and the researcher who originally collected the data.¹⁴ The optic disc margin was drawn by a trained technician.

Visual function measurements

SAP was carried out using the Humphrey Field Analyzer II (Carl Zeiss Meditec, Dublin, California, USA) with the 24-2 testing strategy to evaluate the central visual field. Size III stimuli were presented at 54 locations on a 10-cd/m² uniform background, as a temporal pulse of 200 ms duration. The locations above and below the blind spot were excluded from analysis in this study.

Two versions of CSP (CSP-1 and CSP-2) have been developed and were used in this study. CSP-1 measurements were carried out using a custom-built system based on a commercial visual stimulus generator (VSG2/5, Cambridge Research Systems Ltd.; Cambridge, UK).

The technical specifications of this device are described by Hot et al.¹⁴ The Gabor patches on the VSG had a peak spatial frequency of 0.4 cyc/deg and were presented on a 55-cd/m² background with Weber contrast between 100% and 0.1% in 0.15-log unit steps. The stimuli were presented with a temporal Gaussian envelope of 600 ms (time constant of 100 ms); 68% of energy was within the central 200 ms. CSP-1 determines contrast thresholds for 26 locations over the central visual field, half as many as the 24–2 in order to increase the number of trials and allow greater precision for threshold estimation. An area of the visual field from 23° nasally to 15° temporally and 17° both superiorly and inferiorly was covered by the locations. The central 5° and the area near the typical blind spot were avoided. Although SAP and CSP tested similar regions of the visual field, the stimulus locations were rarely identical.

CSP-2 used the ViSaGe stimulus generator (Cambridge Research Systems Ltd.; Cambridge, UK) to present stimuli similar to those used in CSP-1. Stimuli were presented on a 40-cd/m² background at Weber contrasts from 0.7% to 71%.¹⁶ The CSP-2 stimuli were sine-wave Gabor grating patches with magnification that varied by location: spatial frequency was 0.5 cycle/deg at fixation, decreasing to 0.14 cycle/deg at 21 degrees, based on spatial scale. These sine-wave horizontal gratings were within a two-dimensional Gaussian window whose size was increased as spatial frequency decreased to magnify the stimuli accordingly. The stimuli were presented with 3 cycles of 5 Hz counterphase square-wave flicker. CSP-2 presented stimuli at 57 locations similar to the 24–2 locations as shown in Figure 1 of reference 16 (one blind spot location was excluded in the analyses).

Structure-function pairs

We selected patients who had at least six visits with HRT, SAP, and CSP-1 from the CSP-1 dataset and at least six visits with HRT, SAP and CSP-2 from the CSP-2 dataset. At each visit, the structural and functional measurements were performed within one month. A minimum of 2 months separated each visit. To minimize the impact of learning, we performed our analyses on the last five visits only. We used RA as the structural parameter and MS, for SAP and CSP, as the functional parameter. Three different combinations of structure-function data were assessed: SAP MS and RA, CSP-1 MS and RA, and CSP-2 MS and RA.

RA, SAP and CSP data are each expressed in different units. RA is expressed in mm² (linear unit), the sensitivity values are expressed in logarithmic units for CSP-1 and CSP-2 and in decibels (10 times logarithmic units) for SAP. In order to make direct comparisons between the different measures, we expressed RA and MS in comparable units. We expressed all data in percent of mean normal. For SAP and CSP, we first converted the functional data from their respective logarithmic units into linear units and then calculated MS by averaging the sensitivities at each location. All data (structure and function) were then expressed as percent of mean normal^{14,17} based on mean normal values obtained from independent datasets collected on healthy eyes at IU and SUNY which included 118 (RA), 83 (SAP), 107 (CSP-1), and 52 (CSP-2) healthy eyes.

The DSF Model

The DSF model is illustrated in Figure 1. It uses two descriptors, the position and the velocity vectors.¹⁰ The position vector, computed as the centroid of the paired MS and RA data, is an estimate of the current state of the disease. The velocity vector, computed as the weighted average of the differences in MS and RA values between consecutive visits, provides a description of the direction and the rate at which MS and RA are jointly changing over time. The direction of the vector indicates whether the structural and functional indices are improving or worsening. The length of the velocity vectors provides information about the rate of glaucoma progression, with short vectors indicating slower progression and long vectors indicating faster progression. The DSF model can predict the values expected at the next visit based on the latest position and velocity vectors.

Statistical Analysis

Each combination of structure-function data (SAP MS and RA, CSP-1 MS and RA, and CSP-2 MS and RA) were assessed with the DSF and linear regression. In each, the paired longitudinal RA and MS data from the first 3 visits were used to predict the RA and MS for the 4th visit with the DSF model and with linear regression. Data from the first 4 visits were used for the prediction of RA and MS for the 5th visit. Linear regression was fitted separately to RA and MS over time. The magnitude of the prediction error in percent of mean normal was used to compare the prediction accuracy of the linear regression and DSF models. Prediction error is defined as the square root of the sum of the squared differences for each dimension of the vector of differences between predicted values and actual measurements (L2 norm). We also compared the prediction errors obtained for SAP MS and RA with CSP-1 MS and RA, and with CSP-2 MS and RA within the linear regression and DSF model. This was done to determine whether using CSP instead of SAP could improve the prediction accuracy of the DSF model.

For each model and structure-function pair, we calculated the median PE and the 95% bias-corrected and accelerated (BCa) bootstrap confidence interval.¹⁸ The Wilcoxon signed-rank test was used to determine the presence of significant differences between PEs. All analyses were carried out using R, and SPSS. Because glaucomatous damage often occurs in a localized, rather than in a diffuse pattern, we also performed all analyses in the superior temporal (ST) and inferior temporal (IT) sectors. These sectors have been shown to be the most adversely impacted by glaucoma. The structural and functional data were matched using the Garway-Heath structure-function map.¹⁹

RESULTS

Figure 2 shows the median prediction error and confidence intervals for SAP MS and RA and CSP-1 MS and RA (left panel) and for SAP MS and RA and CSP-2 MS and RA (right panel). The results are shown for global measurements (upper panel), and for measurements in IT sector (middle panel) and ST sector (lower panel). Generally, prediction accuracy was better for the DSF model (triangles) compared to linear regression (circles) when either SAP or CSP were used as the functional test.

For SAP MS and RA in the CSP-1 dataset (Figure 2, upper-left panel), the median prediction errors with DSF (open triangles) for predicting visits 4 and 5 were 6.6% and 8.0% of mean normal, respectively. The corresponding values with CSP-1 MS and RA (closed triangles) were 7.8% (difference of 1.2% compared to SAP MS and RA; $p=0.40$) and 8.6% of mean normal, (difference of 0.6% compared to SAP MS and RA; $p=0.24$) respectively. The DSF performed better in 63% of the eyes when SAP MS was used compared to CSP-1 MS when predicting the 4th visit and in 58% of the eyes for the prediction of the 5th visit. Using the DSF model (for SAP MS and RA in the CSP-1 dataset), the prediction errors were lower by 3.3% ($p = 0.06$) and 0.9% of mean normal ($p = 0.12$) compared to linear regression (open circles) for visits 4 and 5, respectively. The DSF prediction error was smaller compared to linear regression in 74% and 68% of the eyes for the 4th and 5th visits, respectively. Using the DSF model (for CSP-1 MS and RA in the CSP-1 dataset), the prediction errors were lower by 4.0% ($p = 0.02$) and 1.4% of mean normal ($p = 0.31$) compared to linear regression (closed circles) for visits 4 and 5 respectively. The prediction error of the DSF model was smaller in 84% and 58% of the eyes for the prediction of the 4th and 5th visit, respectively.

Similar results were obtained for the CSP-2 dataset (Figure 2, upper-right panel). For SAP MS and RA, the median prediction errors with the DSF model (open triangles) for predicting visits 4 and 5 were 9.0% and 6.3% of mean normal, respectively. The corresponding values with CSP-2 MS and RA (closed triangles) were 9.9% (difference of 0.9% compared to SAP MS and RA; $p=0.82$) and 7.0% of mean normal (difference of 0.7% compared to SAP MS and RA; $p=0.48$) respectively. As observed in the CSP-1 dataset, the DSF model had lower but not significantly different prediction errors for SAP MS than for CSP-2 MS for predicting visit 4 and 5. The DSF performed better in 40% of the eyes when SAP MS was used compared to CSP-2 MS when predicting the 4th visit and in 60% of the eyes for the prediction of the 5th visit. Using the DSF model (for SAP MS and RA in the CSP-2 dataset), the prediction errors were lower by 3.7% ($p = 0.02$) and 4.3% ($p = 0.01$) compared to linear regression (open circles) for visits 4 and 5 respectively. The DSF prediction error was smaller compared to linear regression in 75% and 70% of the eyes for the 4th and 5th visits, respectively. Using the DSF model (for CSP-2 MS and RA), the prediction errors were lower by 5.3% ($p = 0.001$) and 2.4% of mean normal ($p = 0.01$) compared to linear regression (closed circles) for visits 4 and 5 respectively. The DSF model did better in 80% of the eyes for the prediction of the 4th and 5th visits.

In the IT and ST sectors, using CSP MS as the functional test did not improve the prediction accuracy of the DSF model compared with SAP MS. No significant differences were found between SAP MS and CSP-1 MS or between SAP MS and CSP-2 MS for either the DSF model or linear regression. When comparing the DSF model with linear regression, prediction accuracy was similar for these two models in the IT or ST sector. In the IT sector and with the CSP-1 dataset (Figure 2, middle-left panel), the prediction error for SAP MS and RA was smaller for the DSF model than for linear regression by 4.5% ($p=0.003$) for visit 4 and by 1.1% ($p=0.06$) for visit 5. And the prediction error for CSP-1 MS and RA was smaller for the DSF by 4.7% ($p=0.03$) for visit 4 and for linear regression by 0.8% ($p=0.60$) for visit 5. In the CSP-2 dataset (Figure 2, middle-right panel), the prediction error for SAP MS and RA was smaller for the DSF model by 3.6% ($p=0.01$) for visit 4 and for linear

regression by 0.8% ($p=0.09$) for visit 5. The prediction error for CSP-2 MS and RA was smaller for the DSF model by 0.7% ($p=0.004$) for visit 4 and by 3.4% ($p=0.01$) for visit 5. In the ST sector and the CSP-1 dataset (Figure 2, lower-left panel), the prediction error for SAP MS and RA was smaller for the DSF model than for linear regression by 2.0% ($p=0.06$) for visit 4 and 3.3% ($p=0.02$) for visit 5. And the prediction error for CSP-1 MS and RA was smaller for the DSF model by 4.6% ($p=0.02$) and 3.9% ($p=0.02$) for visits 4 and 5, respectively. For the CSP-2 dataset (Figure 2, lower-right panel), the prediction error for SAP MS and RA was smaller for the DSF model by 6.8% ($p=0.003$) and 7.2% ($p=0.002$) for visits 4 and 5, respectively. The prediction error for CSP-2 MS and RA was smaller for the DSF model by 3.5% ($p=0.01$) and 2.7% ($p=0.01$) for visits 4 and 5, respectively.

DISCUSSION

The results of the present study are consistent with those reported in Hu et al¹⁰ in that better prediction accuracy was observed for the DSF model compared to the ordinary linear-regression model in short follow-up series when SAP MS and RA were used. Our results extend Hu et al's finding to the infero- and supero-temporal sectors, which are known to be more sensitive to glaucoma damage. We also show that the DSF model continues to have better prediction accuracy compared to linear regression when CSP is used as the functional test instead of SAP. These results suggest that the DSF model, which uses structural and functional data jointly, may have an advantage over linear regression for early detection of glaucoma progression.

We hypothesized that the lower test-retest variability reported for CSP compared to SAP¹⁴ would further reduce the prediction error of the DSF model. Our results, however, did not support this hypothesis. We found that the errors for global, IT, and ST measurements were not significantly different when either CSP-1 or CSP-2 were used as the functional test compared to SAP. This was observed with both the DSF model and linear regression. Overall, the prediction errors obtained when CSP was used as the functional test were equal to, or slightly greater than the errors obtained when SAP was used as the functional test. The lower test-retest variability for CSP compared to SAP was reported using logarithmic units.¹⁴ It is possible that transforming the data from logarithmic to linear units resulted in both tests in this study having similar test-retest variability. To assess this, we used a different approach to express all data in similar units. For both MS and RA, we computed the logarithm of the patient's value divided by the mean normal value. Figure 3 shows the median prediction errors for global, IT, and ST measurements after converting SAP, CSP-1, and CSP-2 to log units. Compared to when SAP MS was used as the functional parameter, the median prediction error was found to be significantly smaller for CSP-1 MS (for linear regression) and CSP-2 MS (for both the DSF model and linear regression) in the IT sector and for the prediction of visit 4. This result was not, however, found systematically for the global and ST sector and for visit 5. Therefore, expressing the data in the logarithmic scale did not consistently result in smaller errors when CSP was used as the functional test compared to SAP.

The MS for SAP was obtained by averaging the sensitivities at 52 locations, while MS for CSP-1 was obtained by averaging sensitivities at 26 locations. Averaging over more

locations reduces test-retest variability and it is possible that SAP MS effectively had lower test-retest variability compared to CSP-1 in the present study. This could explain the lower error obtained when SAP was used as the functional test compared to CSP-1. To test this possibility, we replicated our analysis after selecting the same number of locations distributed in a similar spatial pattern for both SAP and CSP-1 (see Figure 4). Even when a similar number of locations were used for both tests, no significant differences were observed between the errors for both the DSF model and linear regression. This is consistent with the results we obtained when SAP was used as the functional test compared to CSP-2, where the number of locations was slightly greater for CSP-2 (56 locations) than for SAP (52 locations). Because CSP-1 and CSP-2 differ in ways other than the number of locations they test (placement of the stimuli and varying magnification based on eccentricity), the replication of our analyses with a corresponding number of locations between SAP and CSP-1 was warranted.

Although CSP has lower test-retest variability compared to SAP, other performance characteristics of the test may account for our results. For example, CSP-2 and frequency-doubling technology perimetry have been shown to yield deeper defects than SAP in mild damage, while yielding milder defects in more severely damaged sectors. This has been attributed to lower effects of ganglion cell saturation that occur with CSP compared with SAP,¹⁶ and may result in a faster rate of change for SAP compared to CSP in the early stages of the disease. It is possible that this led to a reduction in the relative prediction accuracy of CSP compared to SAP when each was used as the functional test in the DSF model. The graphs in the left column of Figure 5 show differences in prediction error between SAP and CSP-1 (upper row) and CSP-2 (lower row) for global, IT and ST measurements as a function of depth of defect on SAP calculated as the average over 4 visits. Prediction error tended to be greater for CSP-1 than for SAP for deeper depth of defect (SAP MS between 25% to 75%) and smaller for shallower defects (SAP MS greater than 75%). This is consistent with the observation that CSP defects tend to be deeper than SAP defects in mild damage while SAP defects tend to be deeper later in the disease process.¹⁶ We did not observe, however, the same trend for CSP-2.

In addition to depth of defect, we explored other factors that may explain individual differences in prediction error between CSP-1 and CSP-2 and SAP. Fluctuations in the height of the hill of vision over tests influence model fits and their prediction ability. We would expect lower prediction error for series of visual fields with less fluctuations. These fluctuations are translated into variability in global and sectorial MS, which can be summarized with the standard deviation (SD) of the residuals of their linear-regression fits. The residual SD were obtained from linear regression of MS over time for the first 4 visits. The graphs in the middle column of Figure 5 show differences in prediction error as a function of differences in residual SD of the linear regression fits for MS over time. Another factor that may explain individual differences in prediction error is the signal-to-noise ratio of the fitted models, or equivalently, the ratio of explained to unexplained variance of the model fitted to the series. For linear regression, the signal-to-noise ratio is summarized by the F-statistic. We would expect lower prediction error for greater F-values of the linear regression with MS. We calculated the F-values from linear regression of MS over time for the first 4 visits. The graphs in the right column of Figure 5 show differences in prediction

error as a function of differences in F-values of the linear regression fits for MS over time. There does not seem to be any clear association between differences in residual SD or in F-values and differences in prediction error. Therefore, test-retest variability influenced by fluctuations in the field of vision, and signal-to-noise ratio do not explain individual differences in prediction error for CSP-1 and CSP-2 against SAP.

We used the 2002 Garway-Heath map¹⁹ to determine which visual field locations correspond to the IT and ST sectors. We then averaged these locations without addressing the fact that some visual field locations are situated on the same nerve fiber bundle and therefore enter the optic nerve head at the same location (for example, see the two locations circled in the left panel of Figure 6). These visual field locations do not contribute independent information and ideally, they should have been weighed to reflect this. We therefore used the information provided in Figure 4 of Garway-Heath et al's paper²⁰ to develop weights for each visual field location in the ST and IT sectors. The right panel of Figure 6 shows that the angle at which the bundles of each visual field location enter the optic nerve head are unevenly distributed. We derived the weights by determining the number of optic nerve head degrees represented by each visual field location. When two visual field locations entered the disc at the same degree, the contribution of each of these visual field locations were halved. The MS at each location was multiplied by its corresponding weight and then divided by 45 (the total span in degrees of each of the ST and IT sectors). We reran all SAP MS and RA analyses using the weighted MS and found that the differences in median prediction error for weighted and unweighted SAP MS were small (always lower than 3.2%). Finally, while we used the Garway-Heath map in this study, other maps are also available. We therefore replicated the weighted analyses using the Jansoni map²¹ as implemented in the open source R package visualFields.²² Figure 7 shows the median prediction errors (and 95% confidence intervals) for the IT and ST sectors using weighted MS based on the Garway-Heath map (solid symbols) and the Jansoni map (empty symbols). There were no clear systematic differences between the maps for the IT or ST sectors, among visits and datasets. Both the Garway-Heath and the Jansoni maps use average nerve fiber bundle trajectories to determine the optic nerve head location for each test point in the visual field. But given the large inter-individual anatomical differences²⁰ the test points included in the IT and ST sectors would likely be different if we had access to individualized maps for each eye.

In this study, we did not exclude functional data based on reliability in the primary analysis because of possible differences in reliability indices between CSP and SAP. For example, the stimuli used to test the blind spot in CSP-1 ($\sim 2^\circ$ of visual angle) and CSP-2 ($\sim 3^\circ$) are much larger than the Goldman size III stimulus used in SAP ($\sim 0.4^\circ$). The proportion of fixation losses obtained with CSP-1 and CSP-2 may therefore be greater for CSP than for SAP. To assess the impact of reliability, we performed a secondary analysis which included only data with false-positive rate lower than 33% for all perimetric tests and HRT images with standard deviation of height measure of less than 50 μm . We did not consider false negative errors because these can occur in glaucoma patients who perform reliably.²³ Three patients in the CSP-1 dataset (false-positive errors, $n=2$; HRT height measure standard deviation, $n=1$) and one patient from the CSP-2 dataset (due to HRT height measure standard deviation) did not meet these criteria. An average reduction of approximately 1% of mean

normal in prediction error was observed for CSP-1 compared to SAP for both ordinary linear regression and the DSF model. The greater decrease in prediction error for CSP-1 than for SAP considering reliability resulted in more similar results between their median prediction errors. For CSP-2, the average reduction was less than 0.4% of mean normal in prediction error compared to SAP, which resulted in even smaller median error for CSP-2 than for SAP.

In summary, the findings of this study show similar prediction accuracy when SAP and CSP are used as the functional test within the DSF model. The established lower test-retest variability for CSP-1 and CSP-2 compared to SAP in glaucomatous eyes did not translate into better prediction accuracy within the framework of the DSF model. Signal-to-noise ratio and the impact of test-retest variability on the hill of vision accounted only for a modest percentage of the differences observed in prediction error between SAP and CSP.

ACKNOWLEDGEMENTS

This research was supported by NIH grant EY025756 (LR). Acknowledgment is made to the donors of the National Glaucoma Research, a program of BrightFocus Foundation (LR), for support of this research. This work was also supported in part by a Shaffer grant from the Glaucoma Research Foundation (LR), by an Indiana University - Purdue University Indianapolis DRIVE award (LR), by unrestricted grants from the EyeSight Foundation of Alabama, Birmingham, AL and from Research to Prevent Blindness. The data were collected through NIH grants EY007716 and EY024542 (WHS). The authors wish to acknowledge the contribution of Mitchell W. Dul who oversaw data collection at SUNY.

Funding sources: NIH grants EY025756 (LR), EY007716 (WHS) and EY024542 (WHS). BrightFocus Foundation (LR), Glaucoma Research Foundation (LR), IUPUI DRIVE award (LR), unrestricted grants from the EyeSight Foundation of Alabama, Birmingham, AL and from Research to Prevent Blindness.

REFERENCES

1. Allingham RR, Damji KF, Freedman SF, Moroi SE, Rhee SE, Shields BM. Shields textbook of glaucoma. 6th Edition ed 2012.
2. Naghizadeh F, Garas A, Vargha P, Hollo G. Detection of early glaucomatous progression with different parameters of the RTVue optical coherence tomograph. *J Glaucoma*. 2014;23(4):195–198. [PubMed: 22922666]
3. Artes PH, Chauhan BC. Longitudinal changes in the visual field and optic disc in glaucoma. *Prog Retin Eye Res*. 2005;24(3):333–354. [PubMed: 15708832]
4. Malik R, Swanson WH, Garway-Heath DF. 'Structure-function relationship' in glaucoma: past thinking and current concepts. *Clin Exp Ophthalmol*. 2012;40(4):369–380. [PubMed: 22339936]
5. Hood DC, Kardon RH. A framework for comparing structural and functional measures of glaucomatous damage. *Prog Retin Eye Res*. 2007;26(6):688–710. [PubMed: 17889587]
6. Bowd C, Goldbaum MH. Machine learning classifiers in glaucoma. *Optom Vis Sci*. 2008;85(6):396–405. [PubMed: 18521021]
7. Medeiros FA, Leite MT, Zangwill LM, Weinreb RN. Combining structural and functional measurements to improve detection of glaucoma progression using Bayesian hierarchical models. *Invest Ophthalmol Vis Sci*. 2011;52(8):5794–5803. [PubMed: 21693614]
8. Medeiros FA, Zangwill LM, Girkin CA, Liebmann JM, Weinreb RN. Combining structural and functional measurements to improve estimates of rates of glaucomatous progression. *Am J Ophthalmol*. 2012;153(6):1197–1205 e1191. [PubMed: 22317914]
9. Russell RA, Malik R, Chauhan BC, Crabb DP, Garway-Heath DF. Improved estimates of visual field progression using bayesian linear regression to integrate structural information in patients with ocular hypertension. *Invest Ophthalmol Vis Sci*. 2012;53(6):2760–2769. [PubMed: 22467579]

10. Hu R, Marin-Franch I, Racette L. Prediction accuracy of a novel dynamic structure-function model for glaucoma progression. *Invest Ophthalmol Vis Sci.* 2014;55(12):8086–8094. [PubMed: 25358735]
11. Artes PH, Iwase A, Ohno Y, Kitazawa Y, Chauhan BC. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. *Invest Ophthalmol Vis Sci.* 2002;43(8):2654–2659. [PubMed: 12147599]
12. Turpin A, McKendrick AM. What reduction in standard automated perimetry variability would improve the detection of visual field progression? *Invest Ophthalmol Vis Sci.* 2011;52(6):3237–3245. [PubMed: 21357405]
13. Artes PH, Hutchison DM, Nicolela MT, LeBlanc RP, Chauhan BC. Threshold and variability properties of matrix frequency-doubling technology and standard automated perimetry in glaucoma. *Invest Ophthalmol Vis Sci.* 2005;46(7):2451–2457. [PubMed: 15980235]
14. Hot A, Dul MW, Swanson WH. Development and evaluation of a contrast sensitivity perimetry test for patients with glaucoma. *Invest Ophthalmol Vis Sci.* 2008;49(7):3049–3057. [PubMed: 18378580]
15. Wall M, Woodward KR, Doyle CK, Artes PH. Repeatability of automated perimetry: a comparison between standard automated perimetry with stimulus size III and V, matrix, and motion perimetry. *Invest Ophthalmol Vis Sci.* 2009;50(2):974–979. [PubMed: 18952921]
16. Swanson WH, Malinovsky VE, Dul MW, et al. Contrast sensitivity perimetry and clinical measures of glaucomatous damage. *Optom Vis Sci.* 2014;91(11):1302. [PubMed: 25259758]
17. Shafi A, Swanson WH, Dul MW. Structure and function in patients with glaucomatous defects near fixation. *Optom Vis Sci.* 2011;88(1):130–139. [PubMed: 20935585]
18. Bradley E Introduction to the bootstrap: Monographs on statistics and applied probability. Vol 57 New York: Chapman and Hall; 1993.
19. Garway-Heath DF, Holder GE, Fitzke FW, Hitchings RA. Relationship between electrophysiological, psychophysical, and anatomical measurements in glaucoma. *Invest Ophthalmol Vis Sci.* 2002;43(7):2213–2220. [PubMed: 12091419]
20. Garway-Heath DF, Poinoosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic disc in normal tension glaucoma eyes1. *Ophthalmology.* 2000;107(10):1809–1815. [PubMed: 11013178]
21. Jansonius NM, Nevalainen J, Selig B, et al. A mathematical description of nerve fiber bundle trajectories and their variability in the human retina. *Vision Res.* 2009;49(17):2157–2163. [PubMed: 19539641]
22. Marín-Franch I, Swanson WH. The visualFields package: a tool for analysis and visualization of visual fields. *Journal of vision.* 2013;13(4):10–10.
23. Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability? *Invest Ophthalmol Vis Sci.* 2000;41(8):2201–2204. [PubMed: 10892863]

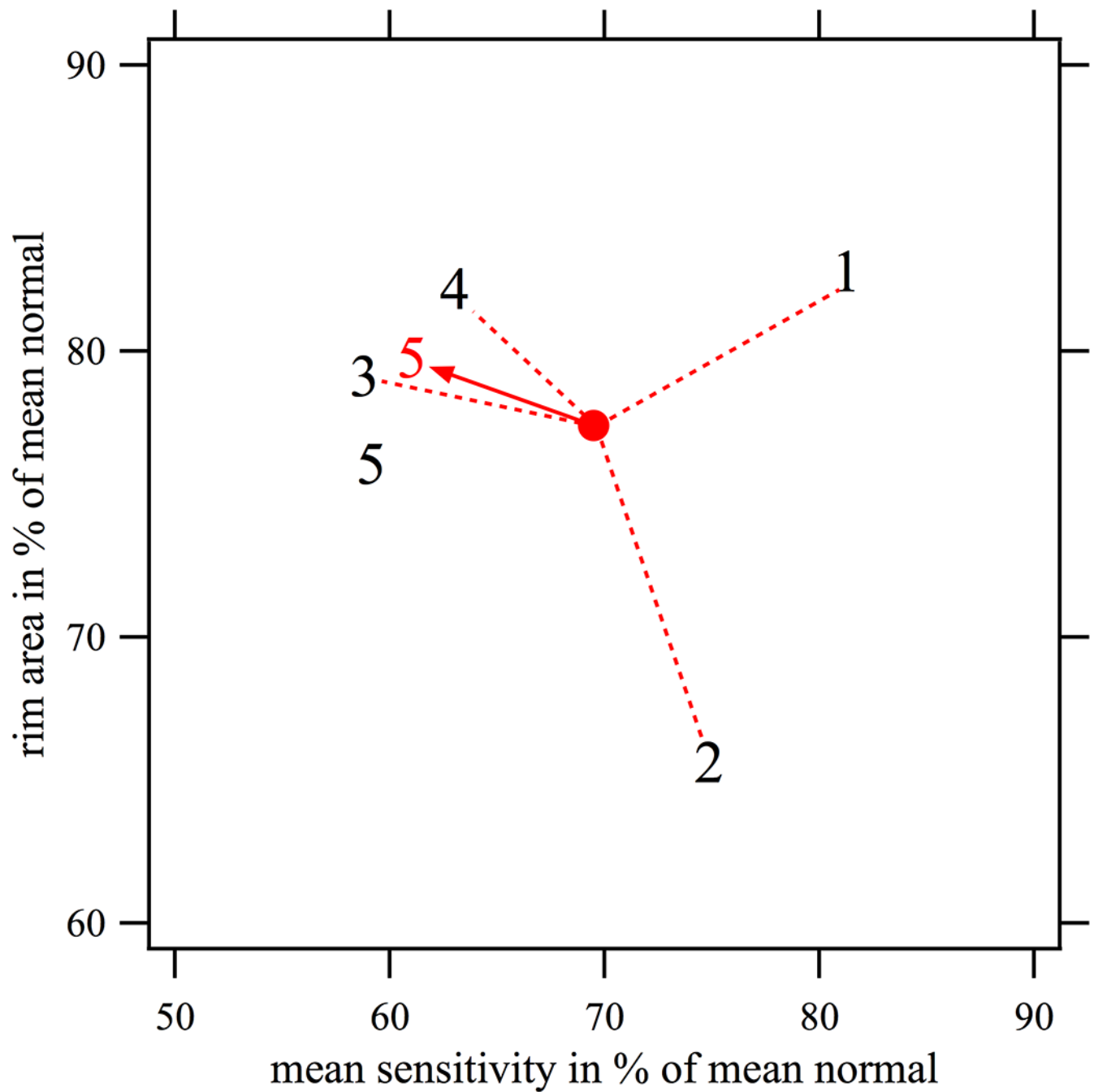


Figure 1.

The position and velocity vectors are represented by the red circle and the red arrow, respectively. The red dashed lines are used to illustrate that both vectors were obtained from the MS and RA values of the first 4 visits (denoted by the numerals 1 to 4). The red numeral 5 represents the predicted values of MS and RA at the 5th visit and is derived from the position and velocity vectors estimated with the DSF model. The prediction error is the distance between the predicted (red numeral 5) and the observed (black numeral 5) values of the 5th visit.

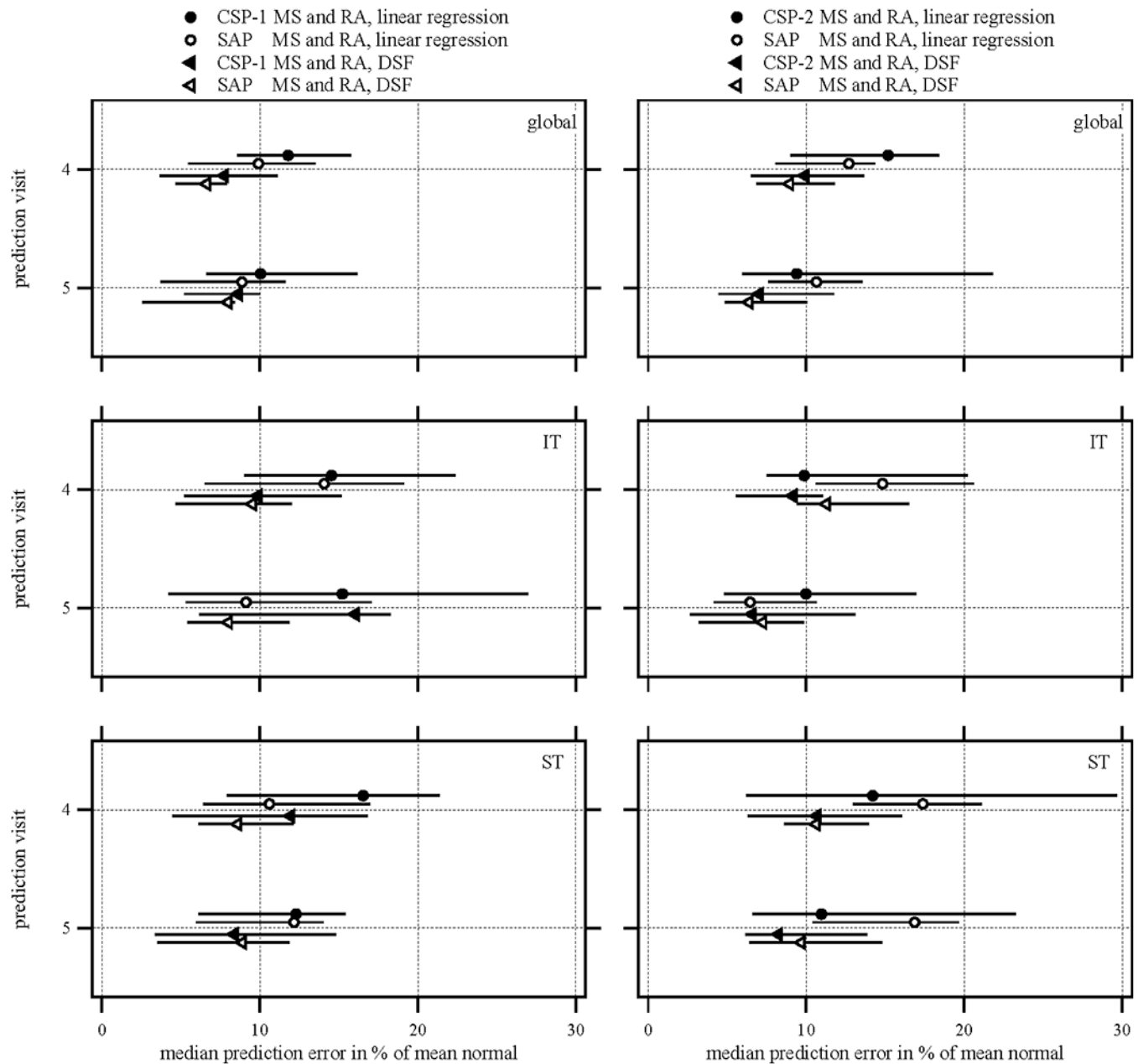


Figure 2.

Median prediction errors (and 95% BCa confidence intervals) for visits 4 and 5 for global SAP MS and RA and CSP-1 MS and RA (n=19) (left panel), and for SAP MS and RA and CSP-2 MS and RA (n=20) (right panel). Results are shown for the DSF model (triangles) and linear regression (circles). The upper panel corresponds to global measurement; the middle and lower panels show the results for IT and ST sectors respectively.

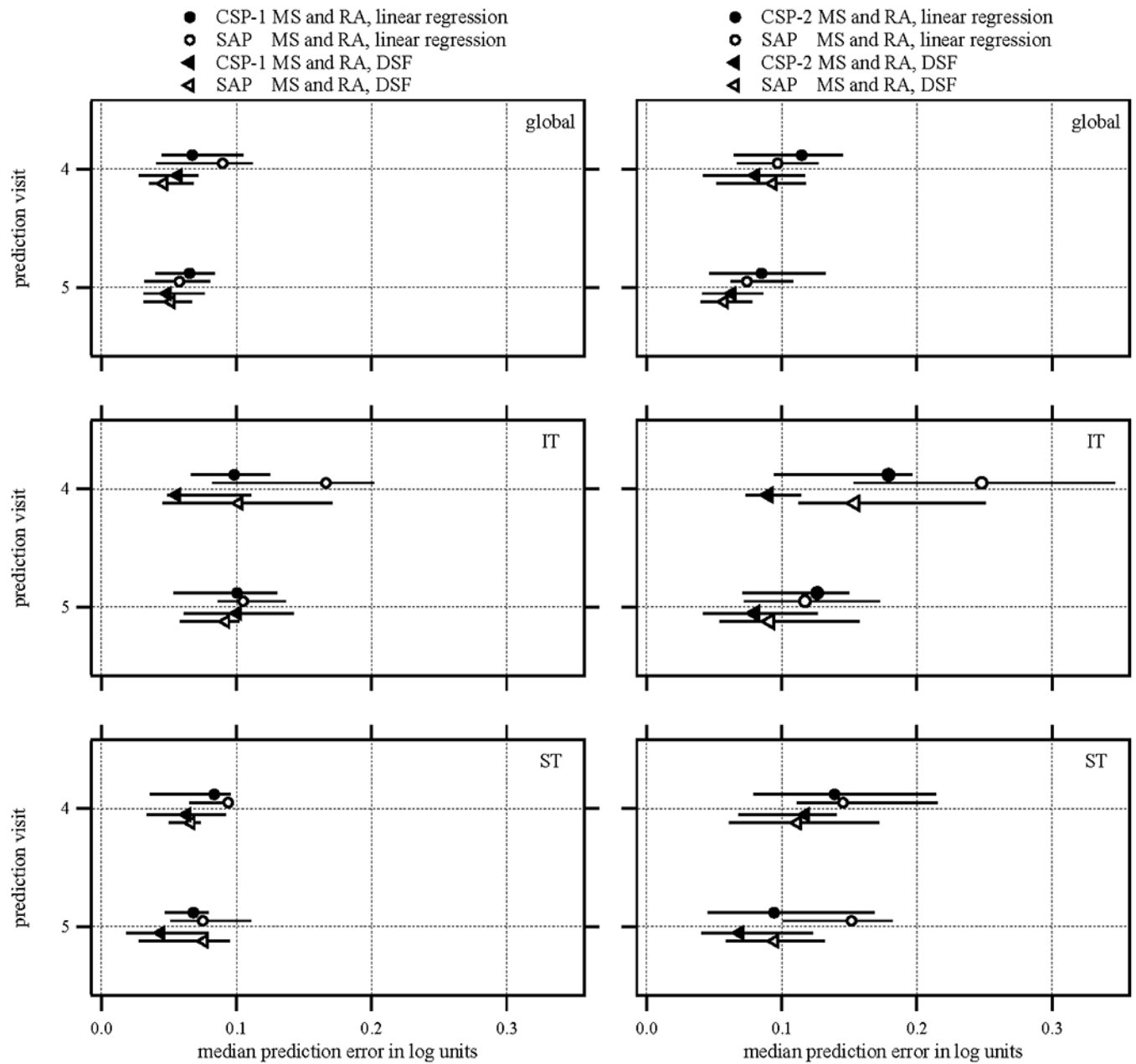


Figure 3.

Median prediction errors of global (upper panel), for the IT sector (middle panel), and the ST sector (lower panel) for % of mean normal based on dB (SAP) and log unit (CSP-1 and CSP-2) scales. Other details as in Figure 2.

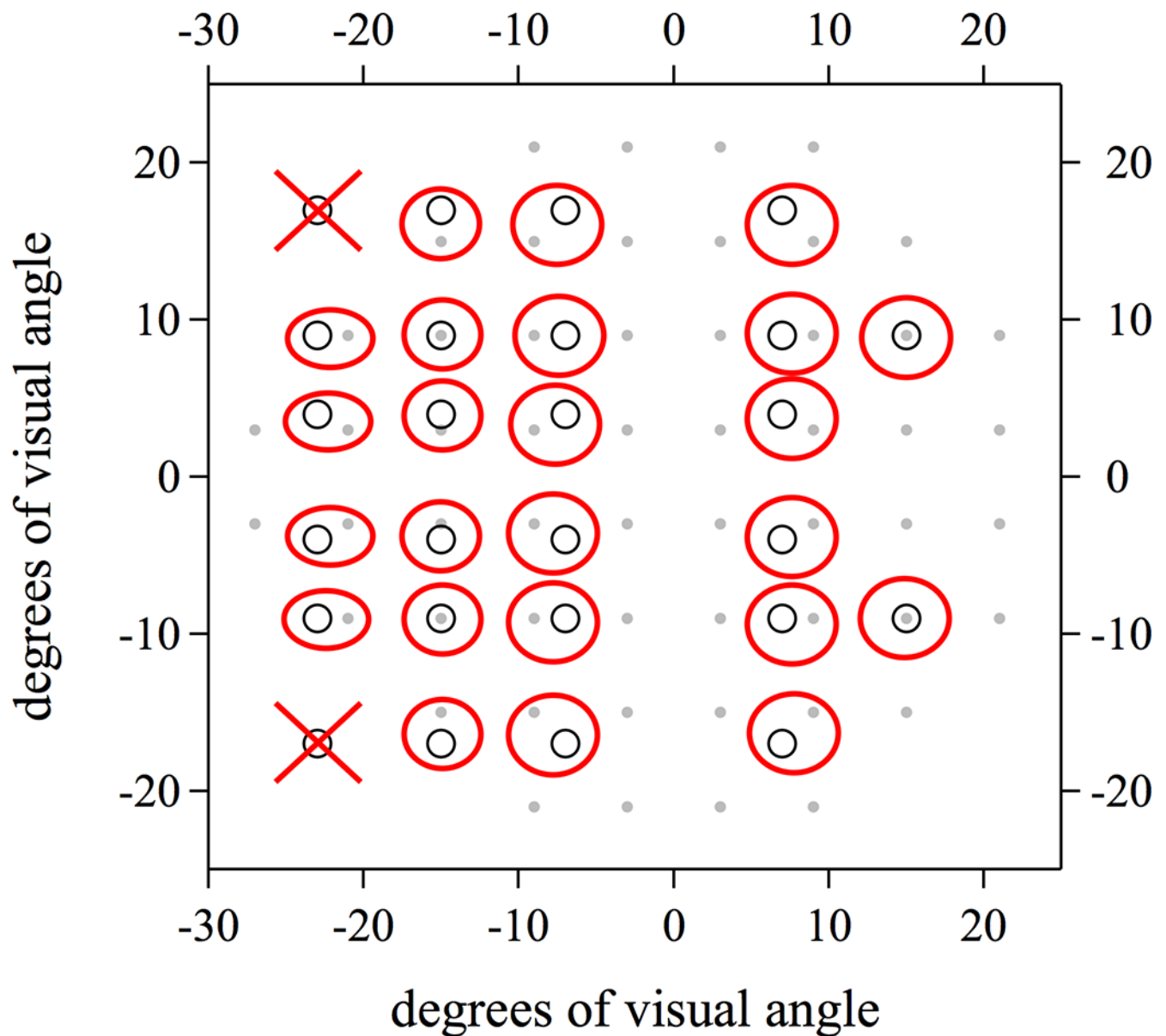


Figure 4.

Pattern of testing location for 24-2 SAP (gray dots) and for CSP-1 (open black circles). The locations selected for both tests are marked with red circles. Thirty locations were removed from the SAP 24-2 test before calculating MS and only two locations (marked with red crosses) were removed from the CSP-1 because they did not overlap with a corresponding SAP test location.

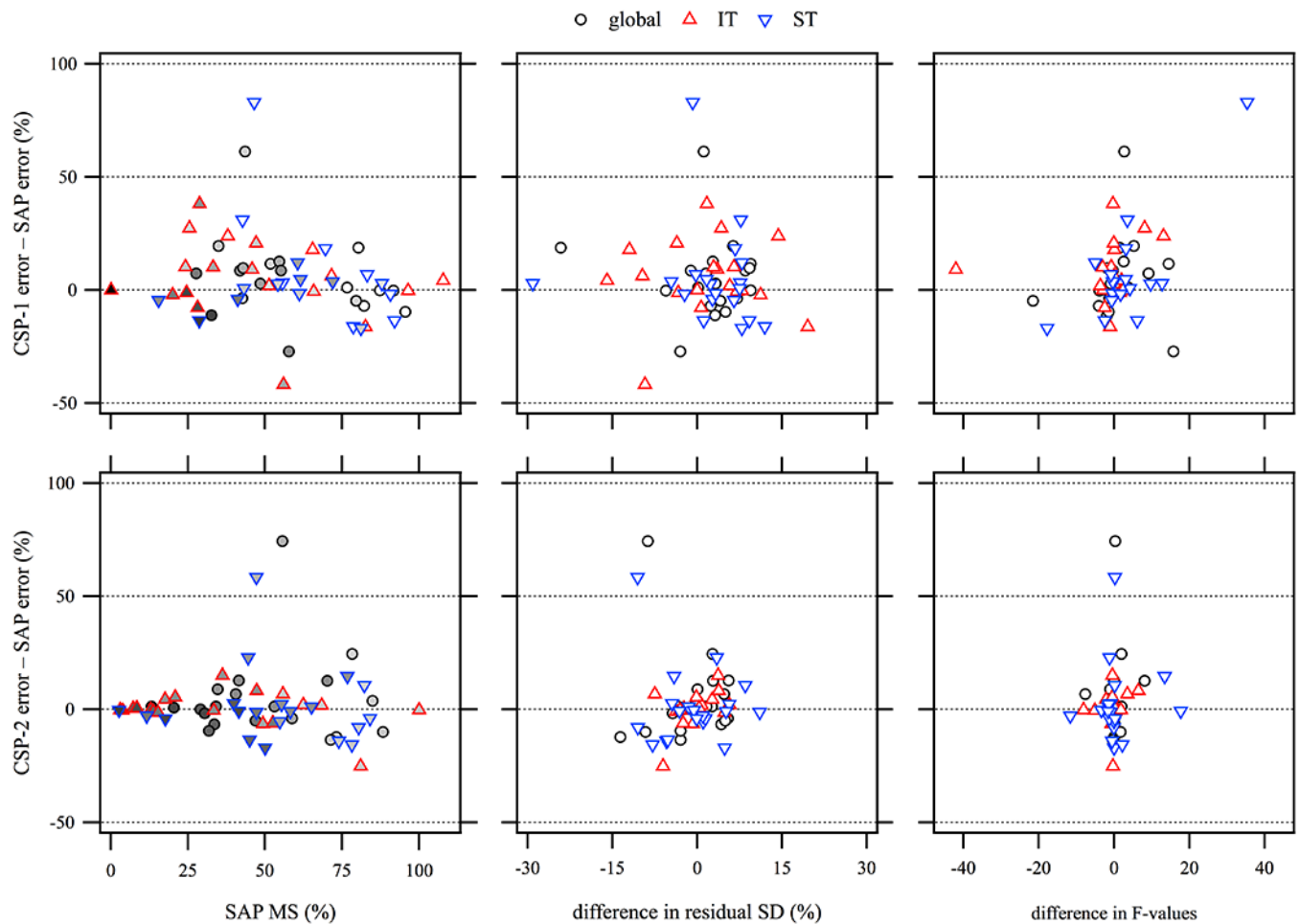


Figure 5.

Differences in prediction error as a function of SAP MS (left column), differences in residual SD (middle column), and differences in F-values (signal-to-noise ratio; right column) for global and IT and ST sectors. CSP-1 or CSP-2 depth of defect in the graphs on the left column are shown with a gray scale with darker shades of gray representing smaller values. The differences in residual SD and F-values in the middle and right columns are between CSP-1 or CSP-2 and SAP. Note: Three points from the upper right graph were removed for clarity since the differences in F-values were disproportionately large at -190 and -637 (ST), and 417 (IT). The differences in prediction error for these three points were -4%, -16%, and -42%. Three points from the lower right graph were also removed for clarity. The differences in F-values of the removed points were 93 and 114 (global), and 143 (ST). The differences in prediction error for these three points were -10%, -2%, and -14%.

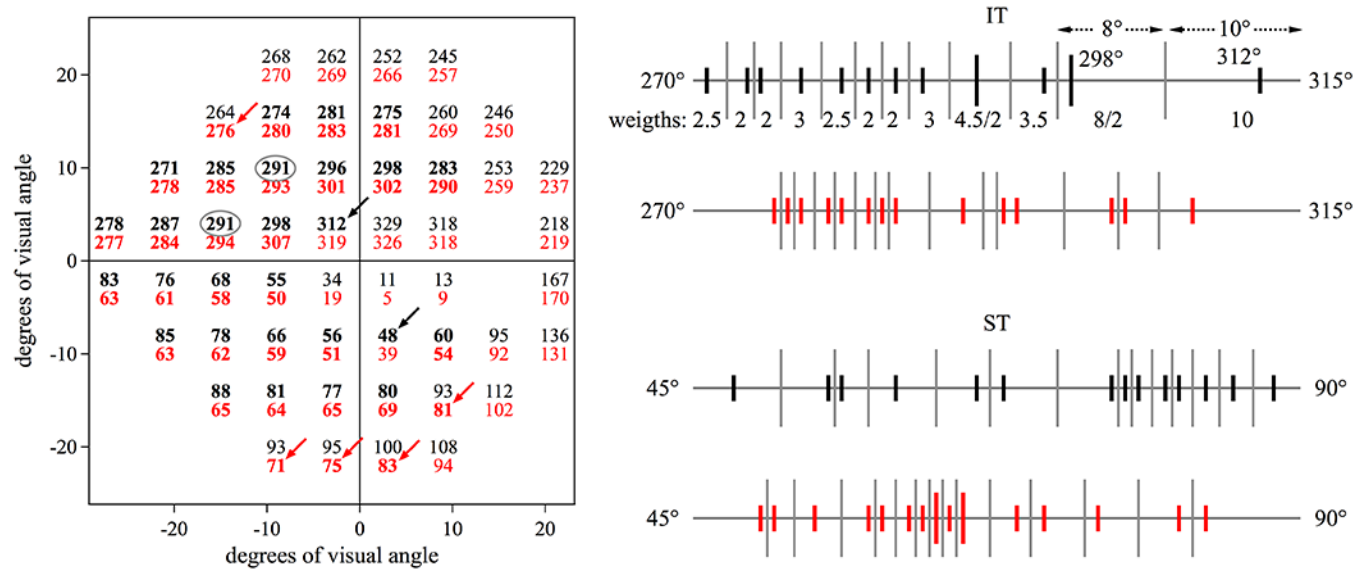


Figure 6.

The angle along the optic nerve head at which the nerve fiber bundles associated with each visual field location is shown in the left panel for the Garway-Heath map (black) and the Jansoni map (red). The bolded locations are those included in the IT and ST sectors. The visual field locations associated with the IT and ST sectors are not identical for the two maps; the arrows indicate visual field locations included in one sector with one map, but excluded from that sector with the other map. The two circles provide an example of two visual field locations served by the same nerve fiber bundle (same angle of entry into the optic nerve head). Weights were applied to the sectoral mean sensitivity, to avoid over-contribution to the MS of visual field locations served by the same bundle. The right panel shows how these weights were derived. Each visual field location located in the IT and ST sectors is mapped based on its angle of entry into the optic nerve head (short bars represent a single visual field location and tall bars represent two visual field locations). The distribution of entry points is uneven and less weight is given to visual field locations that have entry points near to each other. Once each visual field was mapped onto the optic nerve head based on its angle of entry, the midpoints between the angles (vertical light gray bars) were determined. A visual field location received a weight of 10 if the bundles of no other visual field locations entered the optic nerve within 10 degrees of it (as shown for the visual field location whose bundle enters the optic nerve head at 312° in the upper right section of the right panel). When two visual field locations were served by the same bundle, the weight was derived in the same way, but divided by two such that each visual field location was not over-represented.

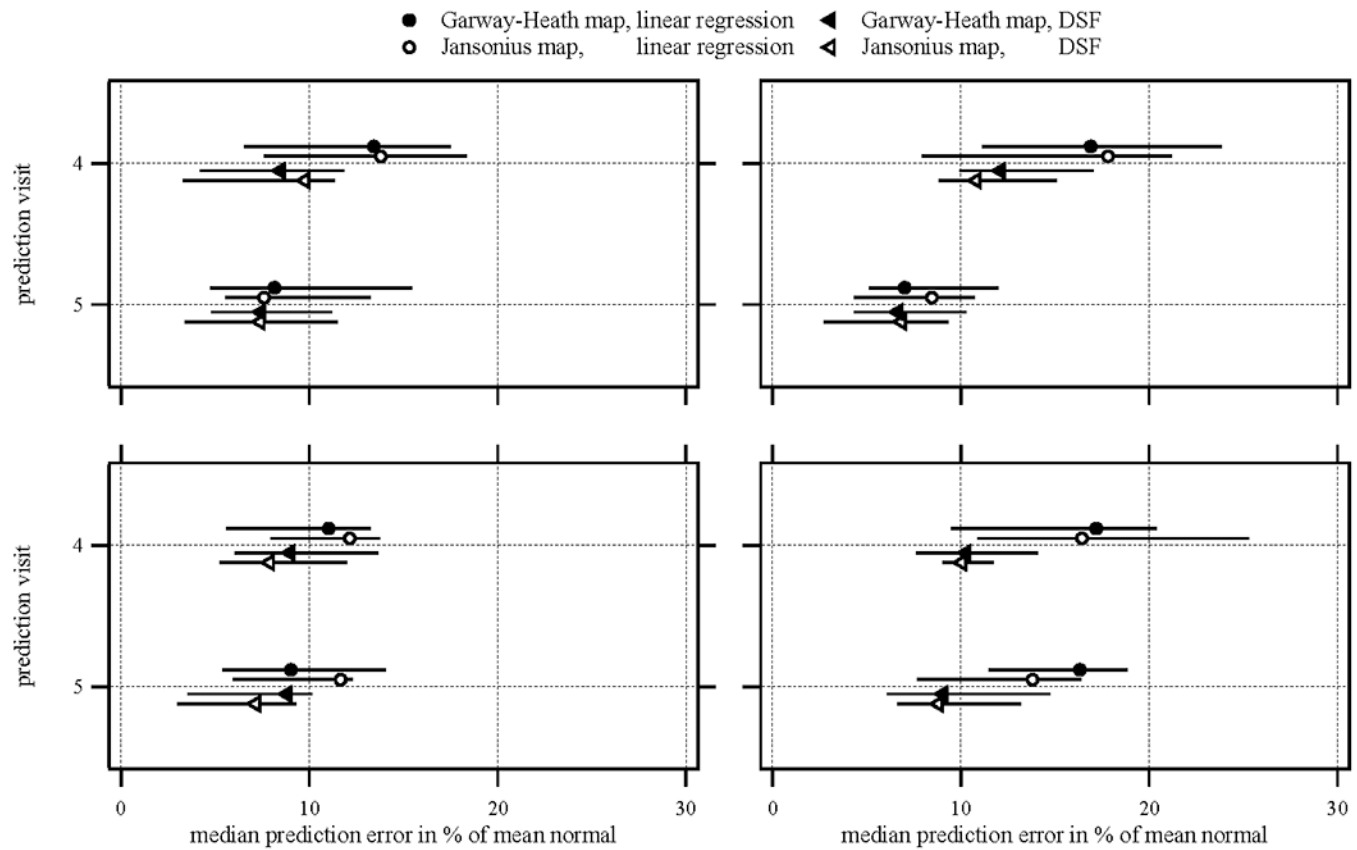


Figure 7. Median prediction errors for the IT (upper panel) and the ST (lower panel) sectors as defined with the Garway-Heath map (solid symbols) and the Jansonius map (empty symbols) for SAP MS and RA in the CSP-1 (left panel) and the CSP-2 (right panel) datasets for linear regression (circles) and the DSF model (triangles). Other details as in Figure 2.